

Faculty of Engineering and Information Technology
University of Technology, Sydney

Actionable Knowledge Discovery: Methodologies and Frameworks

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Dan Luo

June 2009

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

A handwritten signature in black ink, written over a horizontal line. The signature is stylized and appears to be 'D. A. Ho'.

Acknowledgments

I appreciate Professor Chengqi Zhang, my PhD supervisor, for offering me the chance to this study. Without his support, it would be very difficult for me to finish this work.

My second thanks goes to Associate Professor Longbing Cao for his technical suggestions and joint research in my study and to this thesis in particular.

My additional thanks goes to the following schoolfellows for their suggestions and help during my study: Dr. Yanchang Zhao, Dr. Huaifeng Zhang, Dr Jiarui Ni, Mr Yuming Ou, etc., all members in the Data Sciences and Knowledge Discovery Research Lab, as well as all relevant staff and students in the Faculty and the University Graduate School.

I am also grateful to Ms. Li Liu for her continuous support in various management issues.

Finally, I appreciate the support from the APA Scholarship for my research, which contributes to the delivery of this thesis.

Contents

Certificate	i
Acknowledgment	ii
List of Figures	vii
List of Tables	viii
Abstract	ix
Chapter 1 Introduction	1
1.1 Research Motivation and Goals	1
1.1.1 Problem Definition	1
1.1.2 Research Goals	1
1.2 Research Methodology	2
1.3 Research Contributions	4
1.4 Thesis Organization	5
1.5 Summary	6
Chapter 2 Challenges and Prospects	8
2.1 Introduction	8
2.2 KDD Evolution	9
2.3 Challenges and Issues	11
2.3.1 Organizational and Social Factors	12
2.3.2 Human Involvement and Intelligence	13
2.3.3 Domain Knowledge and Intelligence	13
2.3.4 Actionable Knowledge Discovery	14
2.3.5 Decision-Support Knowledge Delivery	15

2.4	Towards Domain-Driven Actionable Knowledge Discovery . . .	16
2.4.1	Problem: Domain-Free vs. Domain-Specific	17
2.4.2	KDD Context: Unconstrained vs. Constrained	18
2.4.3	Interestingness: Technical vs. Business	20
2.4.4	Pattern: Generic vs. Actionable	22
2.4.5	Infrastructure: Automated vs. Human-Mining-Cooperated	24
2.5	Summary	25
Chapter 3	Domain Driven Data Mining Methodologies . . .	26
3.1	Introduction	26
3.2	AKD Fundamental Factors	27
3.2.1	Constrained AKD Environment	27
3.2.2	Catering for Ubiquitous Intelligence	29
3.2.3	Integrating Domain Knowledge	32
3.2.4	Cooperation between Human and KDD Systems	34
3.2.5	Mining In-Depth Patterns	35
3.2.6	Enhancing Knowledge Actionability	37
3.2.7	Closed Loop and Iterative Refinement	38
3.2.8	Interactive and Parallel Mining Support	39
3.2.9	Reference Model	41
3.2.10	Qualitative Research and Questionnaire	42
3.3	D^3M Methodological Framework	43
3.3.1	Theoretical Underpinnings	43
3.3.2	Process Model	44
3.4	Summary	47
Chapter 4	Knowledge Actionability	48
4.1	Introduction	48
4.2	Why Knowledge Actionability?	48
4.3	Knowledge Actionability Framework	50
4.3.1	From Technical Significance to Knowledge Actionability	50
4.3.2	Measuring Knowledge Actionability	54

4.3.3	Narrowing down Interest Gap	56
4.3.4	Developing Business Interestingness	60
4.4	Aggregating Technical and Business Interestingness	62
4.4.1	Specifying Business Interestingness	66
4.5	Summary	66
Chapter 5	Actionable Knowledge Discovery Frameworks . .	68
5.1	Introduction	68
5.2	Why AKD Frameworks	69
5.3	Definition of Actionable Knowledge Discovery	73
5.4	Actionable Knowledge Discovery Frameworks	77
5.4.1	Post Analysis Based AKD: PA-AKD	77
5.4.2	Unified Interestingness Based AKD: UI-AKD	79
5.4.3	Combined Mining Based AKD: CM-AKD	82
5.4.4	Multi-source + combined mining based AKD: MSCM-AKD	85
5.5	Discussions	89
5.6	Summary	92
Chapter 6	Case Studies	93
6.1	Introduction	93
6.2	Case Study 1: Extracting Actionable Trading Strategies	94
6.2.1	What Is Actionable Trading Strategy	94
6.2.2	Constraints on Actionable Trading Strategy Development	98
6.2.3	Methods for Developing Actionable Trading Strategies	102
6.3	Case Study 2: Mining High-Impact Activity Patterns	109
6.3.1	Constructing Activity Sequences	109
6.3.2	Mining Activity Patterns	113
6.3.3	Experimental Results	123
6.4	Summary	129
Chapter 7	Conclusions and Future Work	130
Appendix A	List of Publications	133

CONTENTS

Bibliography	135
------------------------	-----

List of Figures

3.1	Knowledge actionability enhancement	41
3.2	D^3M process model	46
4.1	Fuzzily ranked technical pattern class	63
4.2	Fuzzily ranked business pattern class	64
5.1	Post analysis based AKD (PA-AKD) approach	79
5.2	Unified interestingness based AKD approach	81
5.3	Combined mining based AKD (CM-ADK)	83
5.4	Unsupervised + supervised learning based CM-AKD (USCM-AKD)	85
5.5	Multi-source + combined mining based AKD	87
5.6	Clustering + classification instance	90
6.1	Some results of GA-based trading strategy optimization. . . .	104
6.2	Some results of enhanced trading strategy FR.	107
6.3	Performance comparison: base vs. enhanced trading strategies.	108
6.4	Return on investment of trading strategy-stock pairs.	109
6.5	Activity sequence construction	112

List of Tables

1.1	Key abbreviations.	6
2.1	Data mining development.	10
4.1	Interestingness of data-driven vs. domain-driven KDD.	54
4.2	General interestingness system for AKD	57
4.3	Interest gap between academia and business	58
4.4	Possible inconsistency between technical and business metrics	59
4.5	Relationship between technical and business metrics.	60
6.1	Market organizational factors and impact on rule actionability	100
6.2	Positive and negative impact-oriented activity pattern	114
6.3	Frequent debt-targeted activity patterns in imbalanced set . .	125
6.4	Contrast sequential patterns in target and non-target data . .	126
6.5	Common frequent sequential patterns in separated data	127
6.6	Impact-reversed sequential activity patterns in separated data	128

Abstract

Most data mining algorithms and tools stop at the mining and delivery of patterns satisfying expected technical interestingness. There are often many patterns mined but business people either are not interested in them or do not know what follow-up actions to take to support their business decisions. This issue has seriously affected the widespread employment of advanced data mining techniques in greatly promoting enterprise operational quality and productivity.

In this thesis, a formal and systematic view of actionable knowledge discovery (AKD for short) has been proposed from the system and microeconomy perspectives. AKD is a closed-loop optimization problem-solving process from problem definition, framework/model design to actionable pattern discovery, and to deliver operationalizable business rules that can be seamlessly associated or integrated with business processes and systems. To support AKD, corresponding methodologies, frameworks and tools have been proposed with case studies in the real world to address critical challenges facing the traditional KDD and to cater for crucially important factors surrounding real-life AKD.

First, a comprehensive survey and retrospection on the existing data mining methodologies, issues and challenges in actionable knowledge discovery are reviewed.

Second, a practical data mining methodology: *domain driven data mining* is addressed.

Third, several frameworks have been proposed to support domain driven

actionable knowledge discovery.

Fourth, case studies of domain-driven actionable pattern mining in stock markets and social security data are presented to demonstrate the usefulness and potential of the proposed domain driven actionable knowledge discovery.

In summary, this thesis explores in detail how domain driven actionable knowledge discovery can be effectively and efficiently applied to the discovery and delivery of knowledge satisfying both technical and business concerns as well as to support smart decision-making in the real world. The issues and techniques addressed in this thesis have potential to promote the research on critical KDD challenges, and contribute to the paradigm shift from data-centered and technical significance-oriented hidden pattern mining to domain-driven and balanced actionable knowledge discovery. The proposed methodologies and frameworks are flexible, general and effective to be expanded and applied to mining real-life complex data for actionable knowledge.